MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A.

! ARO 15637.12-M

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A110 766	

①

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
The Construction of Algorithms for the Efficient Estimation of Multivariate Probability Densities	Final Report June 1978–October 1981
	6. PERFORMING ORG. REPORT NUMBER

LEVEL Ⅱ

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
James R. Thompson	DSAAG29-79-C-0180; DAAG29 78 G 0187

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Rice University P.O. Box 1892 Houston, TX 77251	

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709	20 January 1982
	13. NUMBER OF PAGES
	6

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC
ELECTE
FEB 10 1982

B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

unlimited

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DE-CISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

maximum penalized likelihood, representation of densities, onion peel algorithm, intensity function, data based random number generator, nearest neighbor density estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Techniques have been proposed for estimating probability densities nonparametrically. These include penalized maximum likelihood, series, fixed kernel and nearest neighbour procedures. Special attention was given to the two and three dimensional cases. A data-based random number generator for arbitrarily high dimensions was also developed and applied to ballistics data at Aberdeen Proving Ground. A number of biomedical applications were proposed for the related problem of the nonparametric estimation of intervals using cancer data.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE

The Construction of Algorithms for the Efficient
Estimation of Multivariate Probability Densities

Final Report

Author: James R. Thompson

January 20, 1982

U. S. Army Research Office

DAAG-29-79-C-0180

It has been a major objective of this investigation to determine means by which data from an unknown multivariate probability distribution may be most expeditiously summarized, particularly in the exploratory phase of the data analysis process. Unlike the members of exploratory data analysis (EDA) school who attempt to avoid models, our analysis is oriented toward the ultimate goal of modelling the stochastic mechanism at hand. [1] gives a survey of some of the problems outstanding in density estimation.

In [2] we give our algorithm for discrete maximum penalized likelihood estimation of a one dimensional density nonparametrically. This algorithm is now one of the most used for nonparametric probability density estimation-- having been made a part of the IMSL software library (NDMPLE). It should be mentioned that this routine, as a part of the IMSL package, is on disk at practically all DOD computer centers.

In [3], three basic approaches are investigated for estimating two dimensional densities. Computer software is developed for implementing each of the approaches. We note that the two dimensional density estimation problem is actually a three dimensional problem from a representational standpoint. Accordingly, for higher dimensions simple graphing of the density is not possible.

[4] examines the nonparametric estimation of three dimensional densities The "onion peel" algorithm is developed for this task. By this procedure, we first find the modes of the density and then use these as centers for further investigation.
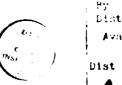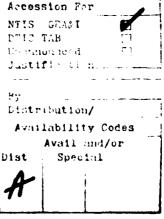
A problem posed by Dr. Malcolm Taylor of the Aberdeen Proving Ground involves large scale simulations in which a multidimensional data set is to be used as a base for a pseudorandom number generator. An algorithm [5] was developed for this prupose which avoids the necessity of estimating the underlying density. It is now operational both at Aberdeen and at Rice.

[6] gives a parametric analysis of breast cancer data in an attempt to model the progress of the disease in a large population of patients. A tentative conclusion is made that in a significant proportion of cases of the disease the five year disease free equivalence of a cure does not appear valid. In [7], we discuss a model-theoretic approach for the optimal scheduling of chemotherapeutic regimes.

The technique of maximum penalized likelihood is employed in [8] for the nonparametric estimation of the intensity of metastasis formation. Contrary to classical neoplastic theory, we discoveredthat the appearance of new metastases is essentially constant in time.

Based on the exploratory data analysis in [8], we developed in [9] a stochastic model for the progress of cancer and used it to analyze a number of data sets from several different kinds of cancer. Our tentative conclusion is that cancer appears to proceed as a systemic disease with the throwing off of metastases relegated to a secondary function. This result gives rise to new speculation as to the causes and treatment of cancer.

REFERENCES

1. Tapia, R. and Thompson, J. [1980]. "Nonparametric density estimation: Where do we go from here?" Proceedings of the 1979 Army Numerical Analysis and Computers Conference, pp. 457-466.

2. Scott, D., Tapia, R. and Thompson, J. [1980]. "Nonparametric probability density estimation by discrete maximum penalized likelihood criteria," Annals of Statistics, V. 8, pp. 820-832.

3. Nezames, D. [1980]. Some Results for Estimating Bivariate Densities Using Kernel, Orthogonal Series and Penalized Likelihood Procedures. Ph.D. Dissertation, Rice University.

4. Fwu, C., Tapia, R. and Thompson, J. [1980]. "The nonparametric estimation of probability densities in ballistics research," to appear in the Proceedings of the Twenty-Sixth Conference of the Design of Experiments in Army Research Development and Testing.

5. Taylor, M. and Thompson, J. [1981]. "A data based random number generator for a multivariate distribution," to appear in the Proceedings of the Twenty-Seventh Conference of the Design of Experiments in Army Research Development and Testing.

6. Brown, B. McBride, C. and Thompson, J., "Is breast cancer cured?" submitted.

7. Thompson, J. [1980]. "A simple model for treatment induced resistance in pest populations and in metastasized tumor systems," in Applications of Information and Control Systems, Lainotes and Tzannes, eds., Reidel Publishing Company, pp. 286-293.

8. Bartoszynski, R., Brown, B., McBride, C. and Thompson, J. [1981]. "Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process," Annals of Statistics, V.9, pp. 1050-1060.

9. Bartoszynski, R., Brown, B. and Thompson, J. [1981]. "Metastatic and systemic factors in neoplastic progression," to appear in Cancer and Statistics, Lecam and Neyman, eds.

Personnel Supported

Chi-Chy Fwu (Ph.D. Student)
Donna Nezames (Ph.D. Student, degree granted 1980)
Richard Tapia
James R. Thompson